Hi I am sehar and now I will explain the slides . this is the EDA process on the one dataset which is the hatecrimes numbers on the basis on most serious violation. So it clearly shows that the most serious violation being observed is Mischief and then Uttering Threats and then Assalult level 1 which is basically intentinally applying force on other person

The other graph is a pie graph shows the same thing but in a pie chart.

Now this is the EDA process on our main dataset we use for our project it is the hate crimes count by type of motivation. This is simple data information , data preview and statistic summary of this data set and along side is the trends of hate crimes over the years it clearly shows that race and ethnicity is on top then comes the religion and the sexual orientation . and so on

This is the comparison of hatecrimes by a bar chart, and identification of outliners by using box plot. Then there is a pairwise scatter plot showing comparison over the years and the last one is the distribution of hate crimes line chart.

Then comes the development phase in which we did predictive analysis on our main dataset

the code aims to build a linear regression model to predict the reported hate crime based on different types of motivation. The model is trained and evaluated, and the predictions are visualized with a scatter plot.

This scatter plot allows you to visually inspect how well the Linear Regression model's predictions align with the actual values. If the blue line closely follows the black dots, it indicates a good fit, suggesting that the model is capturing the relationship between types of motivation and reported hate crimes. If there are deviations, it highlights areas where the model may not be accurately predicting the target variable.

this code trains a Random Forest Regressor model, evaluates its performance on the test set, and visually represents the results through a scatter plot. Random Forest is an ensemble learning method that builds multiple decision trees to improve predictive accuracy and generalization.

Compare the corresponding metrics for both models. Lower values indicate better performance.

This process demonstrates how to extend the use of a trained regression model to make predictions on future data that follows the same structure as the data used during training.

Therefore, based on these evaluation metrics, the Random Forest regression model is more accurate in predicting the total police-reported hate crime based on different types of motivation compared to the linear regression model. The Random Forest model has lower errors, suggesting that it provides a better fit to the data and makes more accurate predictions.

This is just an enhaced viusalisation of make a storymaps on one of our dataset and which is count of hatecrimes in Ottawa .

I will show you

First is intro then dSHBOrd then insights .

**Linear regression model**

This Python code demonstrates a simple linear regression analysis using the scikit-learn library.

the code aims to build a linear regression model to predict the total police-reported hate crime based on different types of motivation. The model is trained and evaluated, and the predictions are visualized with a scatter plot. The evaluation metrics provide insights into how well the model performs on the test set.

This code is similar to the previous one, but instead of using a linear regression model, it employs a Random Forest Regressor to predict the total police-reported hate crime based on different types of motivation.

**Random forest model**

The key difference here is the use of a Random Forest Regressor instead of a linear regression model. Random Forest is an ensemble learning method that builds multiple decision trees and merges their predictions to improve accuracy and control overfitting. The evaluation metrics and the plot help assess how well the Random Forest model performs compared to the linear regression model in predicting hate crime based on different types of motivation.

the Random Forest regression model is more accurate in predicting the total police-reported hate crime based on different types of motivation compared to the linear regression model. The Random Forest model has lower errors, suggesting that it provides a better fit to the data and makes more accurate predictions.

**10 year prediction**

This code is an extension to the previous examples and is designed to make predictions for the total police-reported hate crime for new data from the years 2023 and beyond using the trained linear regression